ELSEVIER

International Journal of Biological Macromolecules
23 (1998) 295–307

INTERNATIONAL JOURNAL OF
Biological
Macromolecules
STRUCTURE, FUNCTION AND INTERACTIONS

Architecture of the sugar binding sites in carbohydrate binding proteins—a computer modeling study

V.S.R. Rao, King Lam, P.K. Qasba *

Structural Glycobiology Section, Laboratory of Experimental and Computational Biology, National Cancer Institute, NCI-FCRDC, Building 469, Room 221, Frederick, MD 21702-1201, USA

Received 24 February 1998; received in revised form 20 June 1998; accepted 2 July 1998

Abstract

Different sugars, Gal, GalNAc and Man were docked at the monosaccharide binding sites of *Erythrina corallodenron* (EcorL), peanut lectin (PNA), *Lathyrus ochrus* (LOLI), and pea lectin (PSL). To study the lectin-carbohydrate interactions, in the complexes, the hydroxymethyl group in Man and Gal favors, *gg* and *gt* conformations respectively, and is the dominant recognition determination. The monosaccharide binding site in lectins that are specific to Gal/GalNAc is wider due to the additional amino acid residues in loop D as compared to that in lectins specific to Man/Glc, and affects the hydrogen bonds of the sugar involving residues from loop D, but not its orientation in the binding site. The invariant amino acid residues Asp from loop A, and Asn and an aromatic residue (Phe or Tyr) in loop C provides the basic architecture to recognize the common features in C4 epimers. The invariant Gly in loop B together with one or two residues in the variable region of loop D/A holds the sugar tightly at both ends. Loss of any one of these hydrogen bonds leads to weak interaction. While the subtle variations in the sequence and conformation of peptide fragment that resulted due to the size and location of gaps present in amino acid sequence in the neighborhood of the sugar binding site of loop D/A seems to discriminate the binding of sugars which differ at C4 atom (galacto and gluco configurations). The variations at loop B are important in discriminating Gal and GalNAc binding. The present study thus provides a structural basis for the observed specificities of legume lectins which uses the same four invariant residues for binding. These studies also bring out the information that is important for the design/engineering of proteins with the desired carbohydrate specificity.

Keywords: Molecular modeling; Legume lectins; Binding specificity

1. Introduction

The discovery of the role of carbohydrates in biological recognition and adhesion process has generated a lot of interest [1]. The studies at the molecular level are of utmost importance in understanding the cellular recognition and other processes in biology. Lectin-carbohydrate interactions have been shown to be involved in a variety of biological processes including mediation of cellular interactions [2]. Lectins of the legume family constitute a large family of homologous proteins and exhibit remarkable divergence in their carbohydrate specificity due to their ability to detect subtle variation in the structure of carbohydrates. Their broad distribution and ease of isolation [3], have been particularly useful in the study of oligosaccharide based molecular recognition process.

Recently, the three dimensional structures of some legume lectins have been determined by X-ray diffraction studies and have been reviewed by Weis and Drickamer [4]. They include con-

* Corrsponding author. Tel. +1 301 8461933; fax: +1 301 8467149; e-mail: qasba@helix.nih.gov.

0141-8130/98/$ - see front matter
PII S0141-8130(98)00056-7

canavalin A (Con A) [5], pea lectin (PSL) [6], favin [7], isolectin I from *Lathyrusochrus* (LOLI) [8], lectin IV from Griffonia-Simplicifolia (GS4) [9], lectin from *Erythrina corallodenron* (EcorL) [3], lentil lectin [10], soybean agglutinin (SBA) [11], and peanut lectin (PNA) [12]. X-ray crystallographic studies have revealed certain characteristic features of the carbohydrate binding site which are independent of their specificities. A specific monosaccharide binds in a narrow cleft on the protein surface which is constituted by residues from four loops A, B, C, and D (Fig. 1a and 1b). These loops are close in the three-dimensional structure, but not in sequence. The invariant residues Asp in loop A, Gly in loop B, Asn in loop C, and an aromatic residue Phe or Tyr also in loop C, interact with the ligand independent of its configuration at C4 atom. From mutational studies, three residues, Asp, Asn, and an aromatic residue have also been identified to be essential for ligand binding irrespective of their specificity [13]. Among the four loops that are involved in the carbohydrate binding, loop D is highly variable in terms of length and sequence. Usually one or two amino acid residues in this loop interact with the ligand, but the residue(s) involved in binding varies from lectin to lectin. The length of loop D in Gal/GalNAc specific proteins is longer as compared to that in Man/Glc specific lectins. It was therefore concluded that the size of this loop is a primary determinant for the carbohydrate specificity of legume lectins [14]. From the mutational studies on EcorL (specific to Gal/GalNAc), Adar and Sharon [13] concluded that of the two residues Ala 218 and Gln 219 in loop D, implicated in carbohydrate binding from X-ray crystallographic studies, the latter is not involved in binding in solution. Furthermore, they concluded that it is not due to Ala 218 in loop D that Man/Glc doesn't bind to EcorL, an assumption that was based on X-ray crystallographic and modeling studies [3]. Hence it is not clear what determines the specificity among legume lectins. Therefore, an attempt has been made to model the binding of Gal, GalNAc and Man with EcorL and PNA, and Man and Gal with LOLI and PSL to determine the structural basis of the sugar binding specificities of legume lectins. Such information helps to understand the architecture of the sugar binding site in legume lectins, which uses the same invariant residues to bind and yet maintain their specificity. This information is also useful to design/engineer proteins with the desired sugar specificity.

2. Results and discussion

Superposition of sugar binding sites of EcorL, PNA, LOLI, and PSL (Fig. 2) show that the invariant amino acids Asp 89, Gly 107, and Asn 133 and the aromatic residue Phe 131 or Tyr, in A, B, and C loops of the lectins as seen in Fig. 2, nearly occupy identical positions independent of their sugar specificity. The residue numbering correspond to the EcorL sequence. Hence, in docking studies, each sugar was initially placed at this site and the energy of the complex was minimized by the procedure described in the methodology section to obtain the best possible orientation of the ligand.

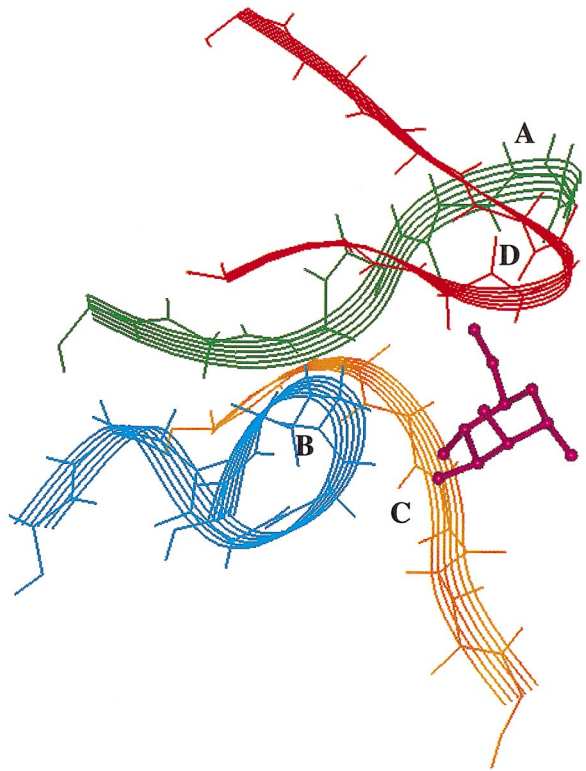
2.1. Complexes of EcorL with β-D-Gal, β-D-GalNAc and α-D-Man

The stereo views of the complexes of EcorL with β-D-Gal, and β-D-GalNAc are displayed in Fig. 3. The possible hydrogen bonds between the protein and the sugar are listed in Table 1.

In EcorL-Gal complex Asp 89 in loop A forms hydrogen bonds with C3–OH and C4–OH hydroxyls of the sugar. C3–OH hydroxyl also forms hydrogen bonds with the backbone NH of Gly 107 in loop B and with side chain of Asn 133 in loop C (Table 1). Asn 133 also forms a hydrogen bond with C2–OH hydroxyl group. From loop D, Ala 218 and Gln 219 form hydrogen bonds with C4–OH and C6–OH hydroxyl groups, respectively. Phe 131 from loop C faces the edge of the sugar having atoms C3–H, C4–H, C5–H, and C6–H2 leading to hydrophobic interactions (Fig. 3). Tyr 106 which has been implicated in sugar binding from X-ray crystallographic studies on lactose complex is placed far away from sugar and is not involved in any interactions with sugar (Fig. 3), which is consistent with the recent mutational studies [13]. The structure of EcorL-lactose complex is known [3], but the EcorL-Gal complex structure has not been reported, except for the hydrogen bond between Asn 133 and O2, all the hydrogen bonds and hydrophobic interactions possible between the Gal residue of lactose and EcorL have been observed in the complex. This suggests that the favored orientation of Gal

```
EcorL: VETISFSFSEFEPGNDNLTLQGASLITQSGVLQLTKINQNGMPAWDSTGR  50
PNA:   AETVSFNFNSFSEGNPAINFQGDVTVLSNGNIQLT-----NLNKVNSVGR  45
LOLI:  TETTSFSITKFGPDQQNLIFQGDGYTT-KERLTLTKAV------RNTVGR  43
PSL:   TETTSFLITKFSPDQQNLIFQGDGYTT-KEKLTLTKAV------KNTVGR  43

EcorL: TLYAKPVHIWDMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKS  100
PNA:   VLYAMPVRIWSSATGNVASFLTSFSFEMKD-IKDYDPADGIIFFIAPEDT  94
LOLI:  ALYSSPIHIWDSKTGNVANFVTSFTFVIDAPNSY-NVADGFTFFIAPVDT  92
PSL:   ALYSSPIHIWDRETGNVANFVTSFTFVINAPNSY-NVADGFTFFIAPVDT  92

EcorL: KPA---QGYGYLGIFNQSKQDNSYQTLGVEFDTFSNPWDPPQ--VPHIGI  145
PNA:   QIPAGSIGGGTLGVSD---TKGAGHFVGVEFDTYSNS-EYNDPPTDHVGI  140
LOLI:  KPQ---TGGGYLGVFNSKDYDKTSQTVAVEFDTFYNTAWDPSNGDRHIGI  139
PSL:   KPQ---TGGGYLGVFNSAEYDKTTETVAVEFDTFYNAAWDPSNRDRHIGI  139

EcorL: DVNSIRSIKTQPFQLDNGQVANVVIKYDASSKLLHAVLVYPSSGAIYTIA  195
PNA:   DVNSVDSVKTVPWNSVSGAVVKVTVIYDSSTKTLSVAVT-NDNGDITTIA  189
LOLI:  DVNSIKSINTKSWALQNGKEANVVIAFNAATNVLTVSLT-YP|ETSYTLN  187
PSL:   DVNSIKSVNTKSWKLQNGEEANVVIAFNAATNVLTVSLT-YP|VTSYTLS  194

EcorL: EIVDVKQVLPEWVDVGLSGATGAQRDAAETHDVYSWSFQASLPE        239
PNA:   QVVDLKAKLPERVKFGFSASG--SLGGRQIHLIRSWSFTSTLITT      232
LOLI:  EVVPLKEFVPEWVRIGFSATTGAE---FAAHEVLSWYFHSELA        227
PSL:   DVVSLKDVVPEWVRIGFSATTGAE---YAAHEVLSWSFHSELSGT      236
```

(a)



(b)

Fig. 1. (a) The structure-based alignment of the sequences in Ecorl, PNA, LOLI and PSL. The loop A, B, C and D is represented as green, cyan, orange, and red, respectively. (b) The sugar binding cleft of LOLI-Gal complex, the loops A, B, C and D indicate the loops involved in the sugar binding.

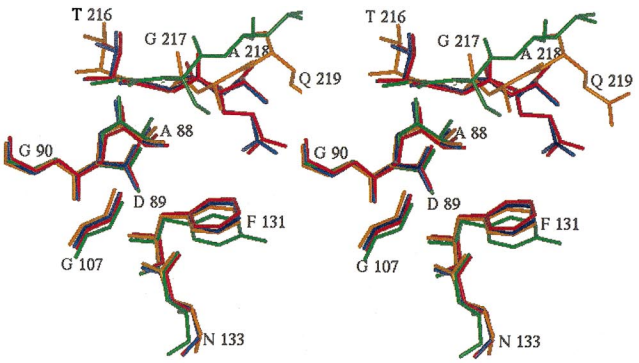


Fig. 2

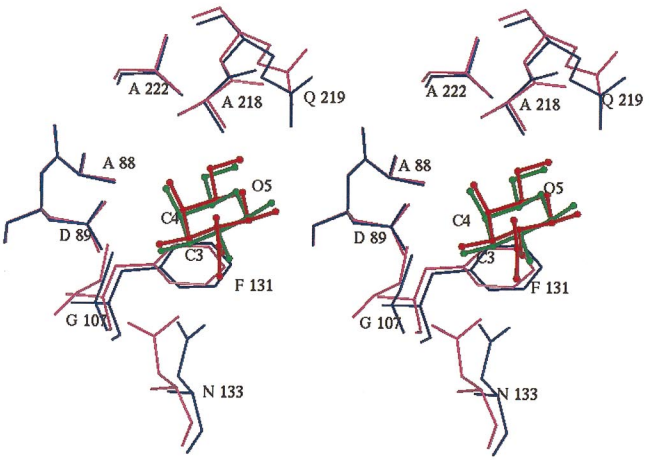


Fig. 3

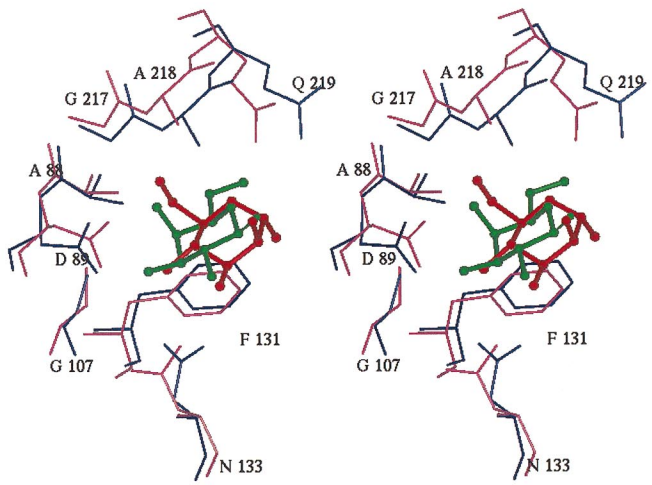


Fig. 4

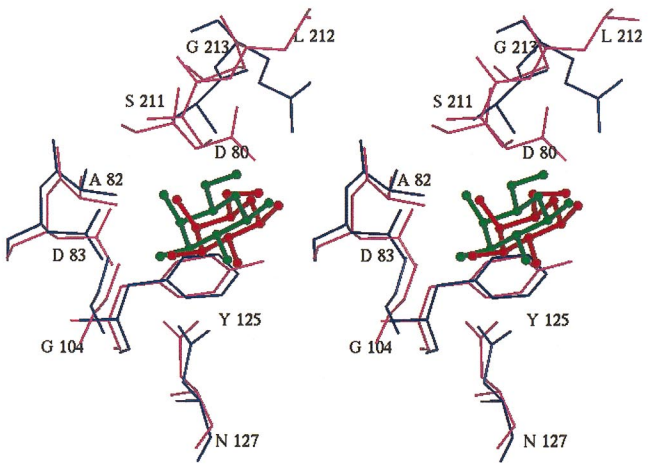


Fig. 5

Fig. 2. The stereoview of the superimposition of the amino acid residues involved in the sugar binding. The amino acid sequence numbering corresponds to EcorL. EcorL residues are shown as orange, LOLI as blue, PNA as green, and PSL as red.
Fig. 3. The superimposition of the carbohydrate binding region of Ecorl in complex with sugars; EcorL (blue)-Gal (green) and EcorL (pink)-GalNAc (red).
Fig. 4. The superimposition of the carbohydrate binding region of Ecorl in complex with sugars; EcorL (blue)-Gal (green) and EcorL (pink)-Man (red).
Fig. 5. The superimposition of the carbohydrate binding region of Ecorl and PNA in complex with sugars; EcorL (blue)-Gal (green) and PNA (pink)-Gal (red).

residue in lactose-EcorL complex is not significantly influenced by Glc residue of lactose.

All the possible hydrogen bonds that are observed in EcorL-Gal complex are also possible in the modeled EcorL-GalNAc complex except the one between Asn 133 and O2. The Phe 131 and Ala 88 are placed below the pyranose ring and close to C6-H2 group, respectively, leading to hydrophobic interactions (Fig. 3). The *N*-acetamido group at C2 atom of GalNAc is not involved in any interactions with the lectin. In EcorL-Gal and EcorL-GalNAc complexes, the CH_2OH group of the sugar favors a *gt* conformation.

In the modeled EcorL-Man complex, the orientation and position of the sugar ring is different than in the EcorL-Gal complex (Fig. 4). The pyranose ring rotates in the binding site so that its hydroxyl groups C6–OH and C4–OH form hydrogen bonds with Asp 89. Asn 133 forms hydrogen bonds with C4–OH and C3–OH hydroxyl groups. Ala 218 and Gln 219 from loop D form hydrogen bonds with C6–OH hydroxyl group and O-5 atom, respectively. Phe 131 is placed below

Table 1
Comparison of the possible hydrogen bonding between monosaccharide (Gal, GalNAc and Man) in the structure EcorL-sugar complexes

Group	Gal	GalNAc	Man
O2	HD22 Asn 133 (2.4; 140)		
O2H			
O3	HN Gly 107 (2.1; 130)	HN Gly 107 (2.1; 131)	
	HD22 Asn 133 (2.1; 139)	HD22 Asn 133 (2.0; 170)	
O3H	OD1 Asp 89 (1.8; 167)	OD1 Asp 89 (1.9; 171)	ND2 Asn 133 (2.2; 142)
O4	HN Ala 218 (2.1; 157)	HN Ala 218 (1.9; 156)	HD22 Asn 133 (2.1; 149)
O4H	OD2 Asp 89 (1.8; 155)	OD2 Asp 89 (1.8; 155)	OD1 Asp 89 (2.0; 139)
			OD2 Asp 89 (2.0; 152)
O5			HE21 Gln 219 (2.3; 122)
O6	HE21 Gln 219 (1.8; 167)	HE21 Gln 219 (2.2; 155)	HN Ala 218 (1.8; 165)
O6H			OD2 Asp 89 (1.7; 161)
N			
NH			
O7			
Docking			
Score[a]	468	509	453
ΔE[b] (kcal/mol)	−30.56	−37.01	−25.93
I_{50} inhibition[c]	5.70 ± 0.52	2.13 ± 0.85	—

In parentheses the distance between the hydrogen and the acceptor atom is given in *Å*, followed by the angle between the hydrogen donor atom, the hydrogen and the acceptor atom in (°). The 'OD', 'ND' and 'HE' refer to '*O* delta', '*N* delta' and '*H* epsilon', respectively.

[a] The data was measured as the docking of the sugar molecule in the combination site of protein which includes some water molecules observed in the crystal structure of EcorL-lactose complex.

[b] $\Delta E = E_{(P.L)} - E_{(P)} - E_{(L)}$: $E_{(P.L)}$ represents the potential energy of the complex, $E_{(P)}$ and $E_{(L)}$ represent the potential energy for protein and ligand, respectively.

I_{50}, concentration (mM) required for 50% inhibition of the binding of the lectin to asialofetuin [19].

the ring leading to hydrophobic interactions. The hydrogen bond between the sugar and the NH of Gly 107 (loop B), the residue which is generally observed in all the legume lectins, is not possible and this may lead to weak binding. These results suggest that Ala 218 is not the one that is responsible for the inability of EcorL to interact to Man. This is in disagreement with the conclusions drawn from the X-ray crystallographic and modeling studies of Shaanan et al. [3], but supports the conclusions drawn from mutational studies. In this mode of binding, CH_2OH group favors a *gg* conformation. Docking score and binding energy, which are the measure of the formation of the complex, are high for GalNAc complex and low for Gal complex which explains the higher affinity of EcorL to GalNAc [19].

2.2. Complexes of PNA with Gal, GalNAc and Man

Peanut lectin (PNA) binds specifically to Gal. The three dimensional structure of PNA with lactose was determined at 2.25 Å resolution [12]. However, the structure of complexes of PNA with Gal, GalNAc, or Man have not been reported. In the modeled PNA-Gal complex, as in EcorL-Gal complex, the most important hydrogen bonds between Gal and Asp 83 from loop A, Gly 104 from loop B, and Asn 127 from loop C have been retained (Table 1, Table 2). The Tyr 125 stacks against the hydrophobic patch (C–4H, C–5H and C6–H2). In the sequence alignment, there is a gap in loop D in PNA corresponding to Ala 218 in EcorL (Fig. 1a), which causes conformational difference at this region. Hence the side chain of the next residue Ser 211 interacts with the sugar rather than the backbone of the peptide. The Ser 211 from loop D forms hydrogen bonds with the ring oxygen (O–5 atom) and C4–OH hydroxyl group. The C6–OH hydroxyl group forms a hydrogen bond with Asp 80 in loop A, unlike in EcorL where it forms a hydrogen bond with Gln 219 in loop D. Except for the hydrogen bond between Asn 127 and the hydroxyl group at C-2 atom of the sugar, the rest of the interactions predicted

Table 2
Comparison of the possible hydrogen bonding between monosaccharide (Gal, GalNAc and Man) in the structure PNA-sugar complexes

Group	Gal	GalNAc	Man
O2	HD22 Asn 127 (2.1; 126)		
O2H			
O3	HN Gly 104 (2.4; 121)	HN Gly 104 (2.3; 139)	HN Gly 104 (2.0; 160)
	HD22 Asn 127 (2.1; 129)	HD22 Asn 127 (2.1; 138)	HD22 Asn 127 (2.4; 145)
O3H	OD1 Asp 83 (1.8; 144)	OD1 Asp 83 (1.9; 150)	
O4	HG Ser 211 (2.2; 121)	HG Ser 211 (2.0; 136)	HD22 Asn 127 (2.2; 140)
O4H	OD2 Asp 83 (1.8; 157)	OD2 Asp 83 (1.8; 153)	OD1 Asp 83 (1.8; 158)
			OD2 Asp 83 (2.3; 125)
O5	HG Ser 211 (2.1; 149)	HG Ser 211 (2.1; 135)	HG Ser 211 (2.1; 130)
O6	HG Ser 211 (1.8; 157)		
O6H	OD2 Asp 80 (1.7; 172)	OD2 Asp 83 (1.7; 170)	
N			
NH			
O7			
Docking			
Score[a]	575	450	389
ΔE[b] (kcal/mol)	−38.83	−32.28	−27.01

In parentheses the distance between the hydrogen and the acceptor atom is given in Å, followed by the angle between the hydrogen donor atom, the hydrogen and the acceptor atom in (°). The 'OD', 'ND' and 'HG' refer to '*O* delta', '*N* delta' and '*H* gama', respectively.

[a] The data was measured as the docking of the sugar molecule in the combination site of protein which includes some water molecules observed in the crystal structure of PNA-lactose complex.

[b] $\Delta E = E_{(P.L)} - E_{(P)} - E_{(L)}$: $E_{(P.L)}$ represents the potential energy of the complex, $E_{(P)}$ and $E_{(L)}$ represent the potential energy for protein and ligand, respectively.

here are as observed with the Gal moiety of lactose in the crystal structure of PNA-lactose complex [12].

Superposition of regions near the sugar binding site in PNA-Gal and EcorL-Gal complexes (Fig. 5) reveal that Gal favors almost identical position and orientation in both the complexes. The possible hydrogen bonds with the invariant residues remain the same, but with the variable residues in loops A and D it may vary. In EcorL, Ala 218 and Gln 219 (loop D) form hydrogen bonds with C4–OH and C6–OH hydroxyl groups, respectively. In PNA, due to the reasons mentioned earlier the side chain of Ser 211 in loop D forms hydrogen bonds with C4–OH and O-5 atom of galactose. Fig. 5 shows that the side chains of Asp 80 (in PNA) and Gln 219 (in EcorL) are placed on either side of the C6–OH hydroxyl group, and are within the hydrogen-bonding distance. Hence, it is not surprising that C6–OH hydroxyl group forms a hydrogen bond with Asp 80 in loop A of PNA and with Gln 219 in loop D of EcorL, although the two residues are not at equivalent position in their sequence of the two proteins. These results suggest that Gal in PNA forms a hydrogen bond to Asp 80 in loop A due to the lack of a polar residue at the equivalent position in loop D, but not due to shorter loop D as suggested from sequence analysis [14]. This indicates that variations in sequences and conformation of loops A and D account for the differences in the hydrogen bond schemes in PNA-Gal and EcorL-Gal complexes. It is interesting to note that in SBA-Gal complex the C6–OH hydroxyl group forms hydrogen bonds with Arg 85 in loop A and with Asp 215 in loop D [15].

Although GalNAc and Gal bind to EcorL in a similar way, they, however, bind differently to PNA. Hydrogen bonds between Asp 83 and hydroxyl groups at C3 and C4 atoms of the sugar, and also the stacking interactions between Tyr 125 and the edge of the sugar ring remain the same for the two proteins. Hydrogen bond distances with Asn 127 and Gly 104 though possible, are larger than usual. The most significant difference seems to be the loss of hydrogen bond between Asp 80 and C6–OH hydroxyl group and an additional hydrophobic interaction between Ile 101 in loop B

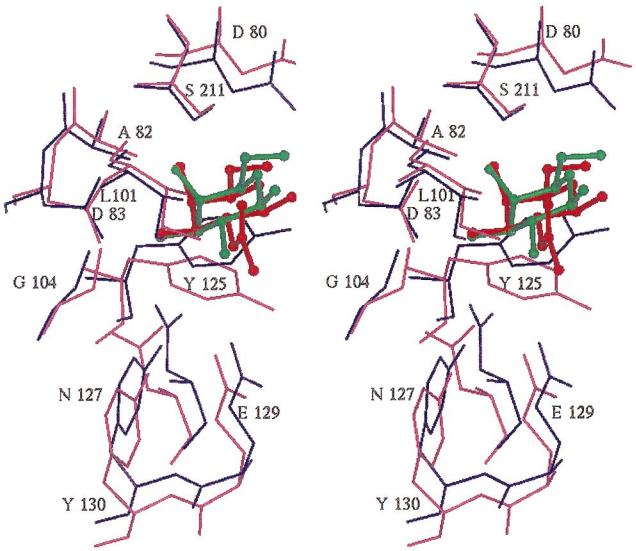


Fig. 6

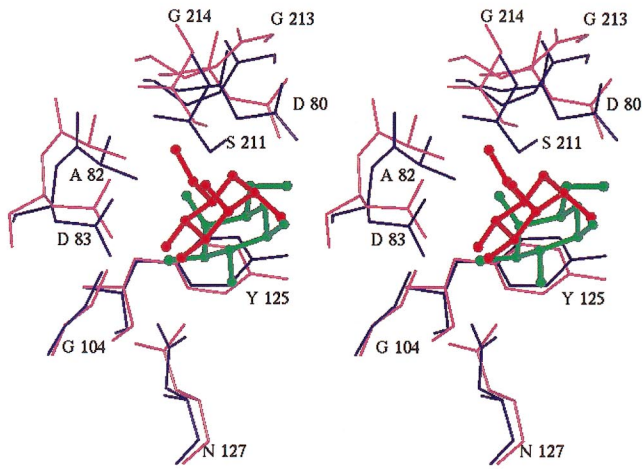


Fig. 7

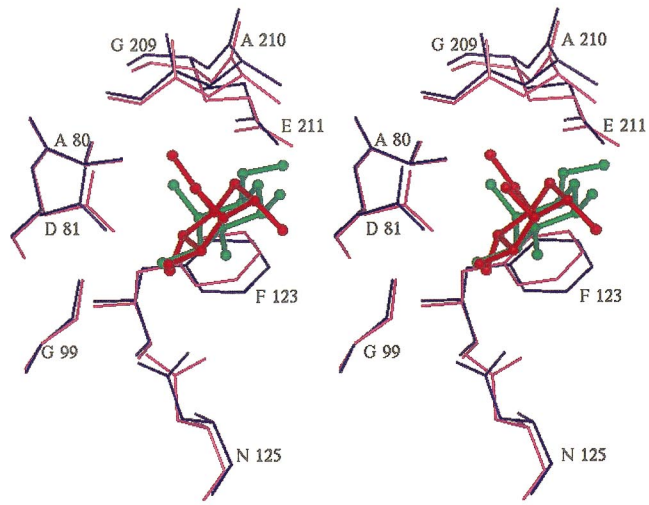


Fig. 8

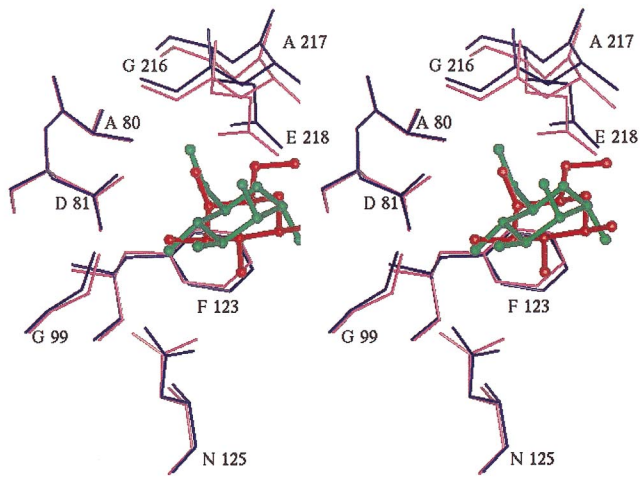


Fig. 9

Fig. 6. The superimposition of the carbohydrate binding region of PNA complex with sugars; PNA (blue)-Gal (green) and PNA (pink)-GalNAc (red).
Fig. 7. The superimposition of the carbohydrate binding region of PNA complex with sugars; PNA (blue)-Gal (green) and PNA (pink)-Man (red).
Fig. 8. The superimposition of the carbohydrate binding region of LOLI complex with sugar; LOLI (blue)-Gal (green) and LOLI (pink)-Man (red).
Fig. 9. The superimposition of the carbohydrate binding region of PSLcomplex with sugars; PSL (blue)-Gal (green) and PSL (pink)-Man (red).

and the acetamide group of GalNAc (Fig. 6). This additional hydrophobic interaction seems to drive the GalNAc to a position where the C3–OH cannot form strong hydrogen bonds with Gly 104 and Asn 217, which also results in the loss of hydrogen bond of C6–OH hydroxyl with the side chain of Asp 80. Our observation is in disagreement with the conclusions drawn [14], based on sequence analysis and interpreted as the inability of PNA to bind GalNAc due to longer loop C as compared to those lectins which bind GalNAc.

In PNA-Man complex, the position and orientation of the sugar is different from that in PNA-Gal complex (Fig. 7). The pyranose ring rotates in the binding site so that Asp 83 in loop A now interacts with C6–OH and C4–OH hydroxyl

groups. Asn 127 in loop C forms a hydrogen bond with C4–OH hydroxyl group. Ser 211 in loop D is placed close to O-5 atom and C6–OH hydroxyl group within the hydrogen bonding distance. Tyr 125 from loop C also stacks against C6–H2 and C5–H groups. But Asp 80 in loop A does not form a hydrogen bond with C6-OH hydroxyl group. Among the three complexes, the docking score and bind energy are high for PNA-Gal complex and low for PNA-GalNAc complex predicting the specificity of PNA for Gal which is in agreement with the experimental observation [16].

2.3. Complexes of LOLI and PSL with Man and Gal—both these lectins are specific for Man/Glc

High resolution X-ray structures of LOLI with α-1-methylmanoside (α-Me-Man) tri- and octosaccharides have been reported [8]. The crystal structure of PSL complexed with methyl 3,6-di-O(α-D-mannopyranosyl)-α-D-mannopyranoside was solved at 2.6 Å resolution [6]. Strong electron density is observed for a single residue indicating that the trisaccharide binds mainly through one of the terminal α-linked mannose residue. The hydrogen bonds, and other interactions between PSL and Man, agree with those predicted by docking studies between PSL and α-MeGlc [17]. In the present study, both Man and Gal are docked at the monosaccharide binding sites of LOLI and PSL as identified by X-ray crystallographic studies. Man forms hydrogen bonds with almost equivalent amino acids in its complexes with LOLI and PSL (Tables 3 and 4). In both the complexes, the aromatic residue stacks against C5–H and C6–H2 below the pyranose ring (Figs. 8 and 9). The predicted hydrogen bonding scheme and the stacking interaction between Man and LOLI or PSL are as observed in the crystal structures of the complexes, between the sugar and these lectins at the monosaccharide binding site except the hydrogen bond between NH of Ala (Ala 210 and 217, respectively, in LOLI and PSL) and C6–OH hydroxyl group.

In the cases described above, Gal binds in an orientation different from that of Man (Figs. 8 and 9). In the monosaccharide binding sites of LOLI or PSL, the pyranose ring of Gal rotates such that its hydroxyl groups at C-3 and C-4 form hydrogen bonds with Asp 81 in loop A. In this mode of binding, the CH_2OH group of Gal moves away from the peptide backbone at loop D, and thus cannot form hydrogen bonds with NH of Ala 210 and Glu 211 as observed in the Man-LOLI complex (Ala 217 and Glu 218 in PSL). Instead, in LOL1 a hydrogen bond between the side chain of Glu 211 in (Glu 218 in PSL) and C6–OH hydroxyl group is possible. The hydroxymethyl group favors *gg* and *gt* conformations in Man and Gal complexes with LOLI and PSL, respectively, just as in EcorL and PNA complexes. The docking score and binding energy are high for the Man complexes compared to their respective Gal com-

Table 3
Comparison of the possible hydrogen bonding between monosaccharide (Gal, GalNAc and Man) in the structure LOLI-sugar complexes

Group	Gal	Man
O2		
O2H		
O3	HN Gly 99 (2.2; 140) HD22 Asn 125 (2.2; 174)	HN Gly 99 (2.2; 170)
O3H	OD1 Asp 81 (1.8; 143)	
O4		HD22 Asn 125 (2.3; 152)
O4H	OD1 Asp 81 (1.9; 150) OD2 Asp 81 (2.2; 137)	OD1 Asp 81 (1.8; 146)
O5		HN Ala 210 (2.2; 146)
O6		HN Ala 210 (2.0; 130) HN Glu 211 (2.0; 167)
O6H	OE2 Glu 211 (1.7; 167)	OD2 Asp 81 (1.7; 171)
Docking Score[a]	436	473
ΔE[b] (kcal/mol)	−23.46	−27.55

In parentheses the distance between the hydrogen and the acceptor atom is given in Å, followed by the angle between the hydrogen donor atom, the hydrogen and the acceptor atom in (°). The 'OD', 'ND' and 'OE' refer to '*O* delta', '*N* delta' and '*O* epsilon', respectively.

[a] The data was measured as the docking of the sugar molecule in the combination site of protein which includes some water molecules observed in the crystal structure of LOLI-α-methyl-D-mannoside complex.

[b] $\Delta E = E_{(P.L)} - E_{(P)} - E_{(L)}$: $E_{(P.L)}$ represents the potential energy of the complex, $E_{(P)}$ and $E_{(L)}$ represent the potential energy for protein and ligand, respectively.

Table 4
Comparison of the possible hydrogen bonding between monosaccharide (Gal, GalNAc and Man) in the structure PSL-sugar complexes

Group	Gal	Man
O2	HD22 Asn 125 (2.4; 148)	
O2H		
O3	HN Gly 99 (2.3; 134)	HN Gly 99 (2.4; 171)
O3H	OD1 Asp 81 (1.9; 162)	
O4	HN Ala 217 (2.6; 137)	HD22 Asn 125 (2.3; 152)
O4H	OD2 Asp 81 (1.8; 159)	OD1 Asp 81 (1.8; 153)
		OD2 Asp 81 (2.3; 125)
O5		HN Ala 217 (2.2; 146)
O6		HN Ala 217 (2.1; 130)
		HN Glu 218 (2.1; 165)
O6H	OE2 Glu 218 (1.7; 166)	OD2 Asp 81 (1.7; 163)
Docking Score[a]	387	452
ΔE[b] (kcal/mol)	−22.68	−26.13

In parentheses the distance between the hydrogen and the acceptor atom is given in Å, followed by the angle between the hydrogen donor atom, the hydrogen and the acceptor atom in (°). The 'OD', 'ND' and 'OE' refer to '*O* delta', '*N* delta' and '*O* epsilon', respectively.

[a] The data was measured as the docking of the sugar molecule in the combination site of protein which includes one water molecule observed in the crystal structure of LOLI-trimannoside complex.

[b] $\Delta E = E_{(P.L)} - E_{(P)} - E_{(L)}$: $E_{(P.L)}$ represents the potential energy of the complex, $E_{(P)}$ and $E_{(L)}$ represent the potential energy for protein and ligand, respectively.

plex suggesting that lectins bind strongly to Man than Gal. This is in agreement with the specificities of LOLI and PSL for Man.

2.4. Importance of the orientation of $-CH_2OH$ group

The pyranose ring of Gal and Man favors the $^4C_1$ conformation both in solution and solid state. However, the orientation of the hydroxyl groups at C-2 and C-4, and also the preferred orientations of the hydroxymethyl group, in the two monosaccharides differ. In Man the hydroxyl group at C-2 atom is in axial orientation, whereas in Gal the hydroxyl group at C-4 atom is in axial orientation. In the solid state, the $-CH_2OH$ group in Gal favors *gt*/*tg*/*gg* conformations in the ratio of 58:34:8 [18]. Same is true in solution also. But in the complexes with the legume lectins it favors only the *gt* conformation. It can be seen with the help of models, that when $-CH_2OH$ group favors a *gt* conformation a hydrophobic patch is created at the edge of Gal constituting the groups C–3H, C–4H, C–5H, and C–6H2. On the contrary, the $-CH_2OH$ group in Man favors *gg*/*gt*/*tg* conformations in the ratio of 60:40:0 in the solid state as well as in solution, but in the complexes with the legume lectins it favors a *gg* conformation to create a hydrophobic patch below the pyranose ring by C–5H and C–6H2 groups. When Man or Gal is docked in the monosaccaride binding site, the hydrophobic patch created by the specific orientation of the $-CH_2OH$ group always places above the aromatic residue (Phe or Tyr) to have favorable hydrophobic interactions. The C6–OH hydroxyl group also forms hydrogen bonds with the lectin. In the case of Man/Glc binding proteins (LOLI and PSL) the *gg* conformation of $-CH_2OH$ group is also essential for the invariant Asp in loop A to form bidentate hydrogen bonds. The length of the fragment that interacts with the lectin is longer when $-CH_2OH$ group favors a *gt* conformation than *gg* conformation. Since it favors *gt* and *gg* conformations in Gal and Man, respectively, the former requires a bigger binding site than the latter. These results suggest that CH_2OH group is a dominant recognition determinant.

2.5. Specificity of legume lectins

When a specific sugar binds at the monosaccharide binding site of lectin, e.g. Gal to EcorL/PNA or Man to LOLI/PSL, it is held by the hydrogen bonds with the amino acids from all the four loops of the lectin. On the other hand, when a less specific monosaccharide is placed in the binding site, e.g. Man in the monosaccharide binding site of EcorL or PNA, one of the hydrogen bonds is weakened or lost (Gly107 in EcorL or Asp80 in PNA) resulting in weak binding (Tables 1 and 2). Similarly when Gal is placed in the monosaccharide binding site of LOLI, $-CH_2OH$ group forms

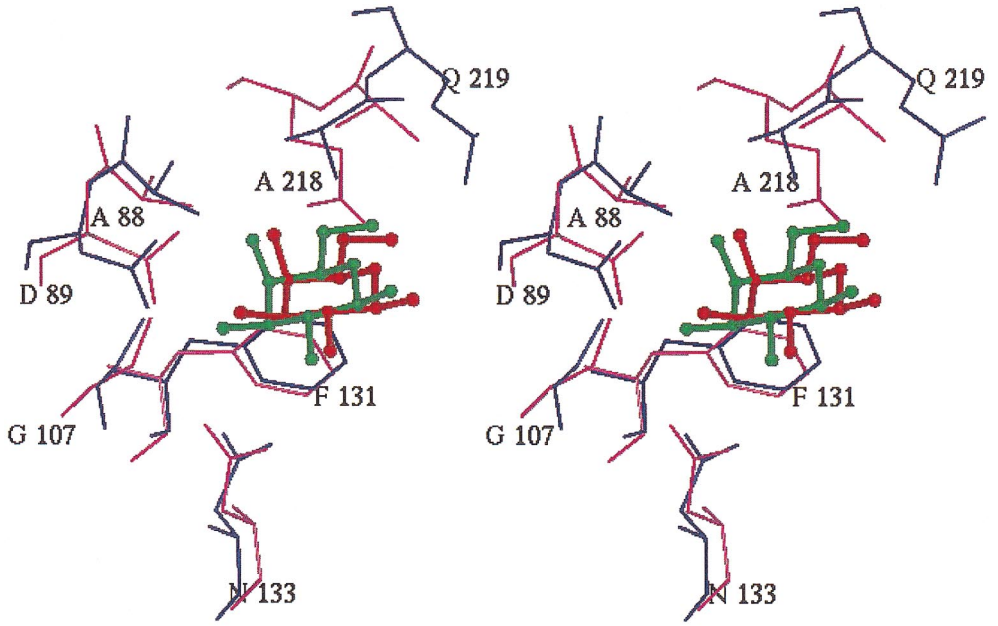


Fig. 10

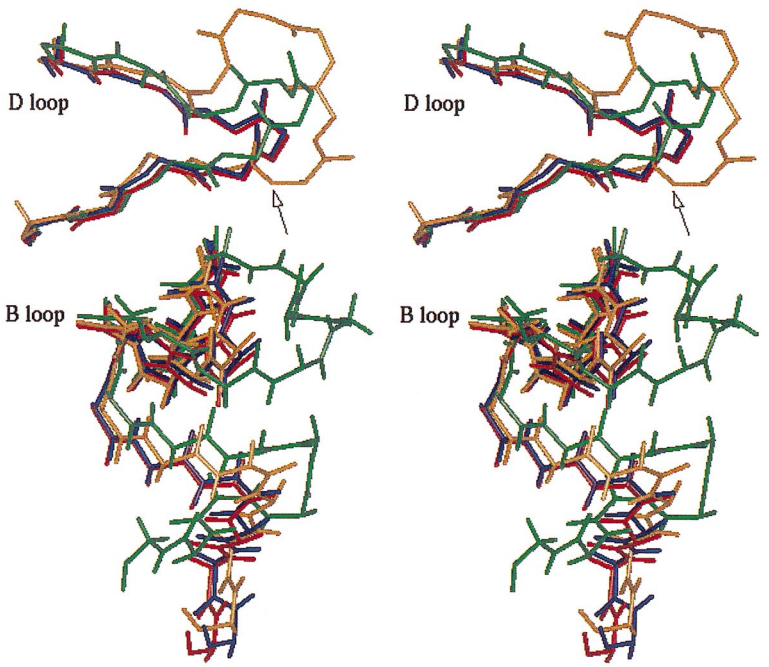


Fig. 11

Fig. 10. The superimposition of the carbohydrate binding region of EcorL and PSL complex with sugars; EcorL (blue)-Gal (green) and PSL (pink)-Gal (red).

Fig. 11. The superimposition of the loop B and D of EcorL, PNA, LOLI and PSL. EcorL is shown in orange, LOLI in blue, PNA in green and PSL in red. The arrow indicates the position of Ala 218 in EcorL sequence numbering.

a hydrogen bond with the side chain of Glu 211 (Glu 218 in PSL) instead of with NH of Ala 210 (Ala 217 in PSL) (Tables 3 and 4). It is also interesting to note that Gal binds almost at the same position and in the same orientation at the sugar binding site (Fig. 10) independent of the specificity of the lectin (EcorL or LOLI/PSL). It also forms similar hydrogen bonds except with one

of the amino acid residues in loop D. This is not surprising since several residues occupy nearly identical positions at the sugar binding site independent of the specificity of the lectin: the residue Asp in loop A that forms bidentate hydrogen bonds with C3–OH and C4–OH hydroxyl groups of Gal, the aromatic residue (Phe or Tyr) from loop C which stacks against the hydrophobic face of the Gal, and Asn also from loop C forms a hydrogen bond with C3–OH hydroxyl group. Superposition of binding site residues (Fig. 1) do reveal that at the loop D small deviations occur near Ala 218 in EcorL (Ala 210 in LOLI and Ala 217 in PSL). Though these differences are small, but they increase the distance between C6–OH hydroxyl and NH Ala 210 in LOLI/PSL to about 3.2 Å leading to weak interaction. On the other hand because of the sharp turn after Ala 210 in LOLI, the deviations in the position of Glu 211 and Gln 219 (in EcorL) are more significant. But the side chains of these residues favor positions on either side of the C6–OH hydroxyl group and within the hydrogen bonding distance. The superposition of EcorL-Man and LOLI/PSL-Man complexes suggest that Man in the binding site of EcorL moves slightly towards loop D to retain the hydrogen bond between NH Ala 218 and C6–OH hydroxyl group, but loses the hydrogen bond with the NH of the invariant Gly107. Due to significant conformational change at Gln 219, the hydrogen bond between C6–OH hydroxyl group and NH Gln 219 (the equivalent of C6–OH and NH Glu 211 in LOLI) is not possible. Instead the side chain of Gln 219 forms a hydrogen bond with O-5 atom of the sugar. It thus seems that subtle variations in the conformation of loop D seems to control the specificity of lectins to C-4 epimers. Of the two lectins, EcorL and PNA, the former binds to both Gal and GalNAc while the latter binds only to Gal. Since both Gal and GalNAc occupy nearly the same position near the D loop (as revealed in docking studies), the differences in their interactions with PNA and EcorL is not due to the difference in the length of the loop D, but due to the presence of Ile in loop B of PNA. Also in SBA-GalNAc complex, *N*-acetamido group of GalNAc interacts with Tyr 107 in loop B and Asn 130 in loop C. It thus seems that the binding of sugar to lectin may also be effected by the variations at loop B, since Asn from loop C instead of interacting with C2–OH or C3–OH now interacts with *N*-acetamindo group.

The variations in the conformation of the peptide fragment observed at loops B or D might have originated from the difference in the size of the sequence and also the location of gaps in these loops as they connect the similar structural elements in these lectins. Structure based sequence alignment of the amino acids of EcorL, PNA, LOLI and PSL shows that gaps are present at all the four loops, but they are more significant at loops B and D (Fig. 1). At loop B, the size of the gap is equal in all the lectins, but it appears at different locations, at least three residues away from the invariant Gly. This seems to affect the conformation of the loop B differently and hence its interaction with the sugar (Fig. 11). On the other hand, in loop D, not only the size of the gap is different in different lectins, but the location of gap leads to drastic difference in the conformation of loop D at the end of the binding edge (after Ala 218 in EcorL numbering) that is involved in hydrogen bonding with the sugar. However, subtle variations have also taken place at Ala 218 in loop D, which affects the hydrogen bonding involved with the backbone—NH of the peptide chain. This perhaps explains the observed difference in hydrogen bonding between sugar and the equivalent amino acid residues of loop D when the same sugar (e.g. Gal) is placed at the sugar binding site in different lectins. It thus seems that the location and size of the gaps seem to influence the conformation of the loop, and hence the specificity of the lectin.

Four of the invariant residues Asp in loop A, Gly in loop B, Asn, and an aromatic residue in loop C occupy nearly identical positions in the three dimensional structures of these lectins. It is thus interesting to note that usually when the specific sugar binds in the monosaccharide binding site of lectin, the sugar is held at both ends by the amino acid residues at loops B and D, except in the case of PNA where the $-CH_2OH$ group is also held tightly by Asp 80 in loop A. These hydrogen bonds are disrupted when a nonspecific sugar is placed at the monosaccharide binding site, but interactions with the invariant residues, Asp in loop A, Asn and the aromatic residue Phe/Tyr in loop C, are retained independent of the configuration of sugar at C-4 atom. Once the sugar is recognized/held by this site, the amino acid residues in loops B and D/A hold the sugar tightly on either side of the binding site. Loss of these

interactions lead to weak bindings. Since Gly in loop B is invariant, variations in conformation and sequence of loop D (A in the case of PNA) seem to control the specificity of these lectins. However, variations at other loops also affects the specificity of the ligand which differ at other carbon atoms, e.g. at C-2 as is found in the docking studies of GalNAc.

In conclusion, in legume lectins, independent of their sugar specificity, the invariant amio acid residues Asp in loop A, Asn and an aromatic residue in loop C provide the necessary architecture to recognize and interact with the sugar having galacto or gluco (manno) configuration at C-4 atom. The invariant residue Gly in loop B together with one/two residues in the highly variable loop D (or loop A in PNA) holds the sugar tightly at either ends of the monosaccharide binding site. Loss of hydrogen bonds at either of these places leads to the formation of weak or no complex. Lastly, specific orientation of the $-CH_2OH$ group, (i.e. *tg* in Gal and *gg* in Man) that directs the hydrophobic patch in the sugar to interact with the aromatic residue and form hydrogen bonds at the binding sites seem to be a dominant recognition determinant. Thus, the present docking studies provide the structural basis for the understanding of lectin carbohydrate interactions. These studies may also help to design/engineer proteins of desired carbohydrate specificity.

2.6. Computational methods

All model building, docking and energy minimization studies were performed on a Silicon Graphics Octane workstation equipped with a stereo graphic facility and Cray J916 Parallel vector supercomputer using InsightII ver. 95.6/Discover ver. 2.97 with AMBER force field (Biosym Technologies, San Diego). The docking calculations were accomplished by Docking module of InsightII. Docking of new ligands into the binding site was carried out by superimposing selected atoms from the new ligand onto the corresponding atoms from the ligand of the complex in crystal structure, and the new complex was then energy—minimized to refine the interaction of new ligand with protein. The monosaccharides (Man, Gal and GalNAc), with $^4C_1$ ring conformation, were initially placed in the combining site of legume lectins in an orientation such that it interacts at least with four of the invariant amino acids Asp, Gly, Asn and Phe (Tyr). The structure of complex was energy- minimized by 100 steps of steepest descent (with cross-terms turned off), followed by conjugate gradient interaction steps until the root mean square deviation (rmsd) was less than 0.01 kcal/mol. During this study, Ca^{2+} and Mn^{2+} were restricted with a distance about 4.5 (Å) and a dielectric constant of 4 ($\varepsilon = 4 \cdot \gamma$) was used throughout the entire study.

The docking score of sugar at the binding sites of the lectins was carried out using the Ligand_Design program interfaced with InsightII. The scoring functions relied on shape complementarity, and take into account both the area of contact and the hydrogen bonding interactions between the ligand and the protein [20]. The binding energy between the ligand and the protein has been considered as the interaction energy (ΔE) from the minimized structures of complexed and separated forms. $\Delta E = E_{(P.L)} - E_{(P)} - E_{(L)}$; $E_{(P.L)}$; $E_{(P.L)}$ represents the potential energy of the sugar-protein complex, $E_{(P)}$ and $E_{(L)}$ represent the potential energy for protein and ligand, respectively.

Acknowledgements

The authors acknowledge the Frederick Biomedical Super computing Center of the Frederick Cancer Research and Development Center, National Cancer Institute for allocation of computing time and staff support.

References

[1] Ashwell G, Harford J. Carbohydrate-specific receptors of the liver. Annu Rev Biochem 1982;51:531–54.
[2] Lasky LA. Selectins: interpreters of cell-specific carbohydrate information during inflammation. Science 1992;258:964–9.
[3] Shaanan B, Lis H, Sharon N. Structure of a legume lectin with an ordered *N*-linked carbohydrate in complex with lactose. Science 1992;254:862–6.
[4] Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. Annu Rev Biochem 1997;65:441–73.
[5] Derewenda Z, Yariv J, Helliwell JR, Kalb AJ, Dodson EJ, Papiz MZ, Wan T, Campbell J. The structure of the saccharide-binding site of concanavalin A. EMBO J 1989;8:2189–93.
[6] Rini JM, Hardman KD, Einspahr H, Suddath FL, Carver JP. X-ray crystal structure of a pea lectin-triman-

noside complex at 2.6 Å resolution. J Biol Chem 1993;268:10126–32.
[7] Reeks GN, Recker JW. Three-dimensional structure of Favin: saccharide binding-cyclic permutation in leguminous lectins. Science 1986;234:1108–11.
[8] Bourne Y, Roussel A, Frey M, Rouge P, Fontecilla-Camps JC, Cambillau C. Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. Proteins 1990;8:365–76.
[9] Delbare LT, Vandonsclaar M, Prasad L, Quail JW, Nijrad PV, Spohr U, Lemieux R. Molecular recognition of human blood group determinant by a plant lectin. Can J Chem 1990;68:1116–21.
[10] Loris R, Lisgarten J, Maes D, Pickersgill R, Korber F, Reynolds C, Wyns L. Two crystal forms of the lentil lectin diffract to high resolution. J Mol Biol 1992;223:579–81.
[11] Dessen A, Gupta D, Sabesan S, Brewer CF, Sacchettini JC. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. Biochemistry 1995;34:4933–42.
[12] Banerjee R, Das K, Ravishankar R, Suguna K, Surolia A, Vijayan M. Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. J Mol Biol 1996;259:281–96.
[13] Adar R, Sharon N. Mutational studies of the amino acid residues in the combining site of *Erythrina corallodendron* lectin. Eur J Biochem 1996;239:668–74.
[14] Sharma V, Surolia A. Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. J Mol Biol 1997;267:433–45.
[15] Rao VSR, Lam K, Qasba PK. Three dimensional structure of the soybean agglutinin-Gal/GalNAc complexes by homology modeling. J Biomol Struct Dyn 1998;16:In press.
[16] Pereira ME, Kabat EA, Lotan R, Sharon N. Immunochemical studies on the specificity of the peanut (*Arachis hypogaea*) agglutinin. Carbohydr Res 1976;51:107–18.
[17] Rao VSR, Biswas M, Mukhocadhyay C, Balaji PV. Computer simulation of protein-carbohydrate complexes: application to L-arabinose binding protein and pea lectin. J Mol Struct 1989;194:203–14.
[18] Marchessault RH, Perez S. Conformations of hydroxymethyl groups in crystalline aldopyranoses. Biopolymer 1989;18:2369–74.
[19] Arango R, Rodriguez-Arango E, Adar R, Belenky D, Loontiens FG, Rozenblatt S, Sharon N. Modification by site-directed mutagenesis of the specificity of *Erythrina corallodendron* lectin for galactose derivatives with bulky substituents at C-2. FEBS Lett 1993;330:133–6.
[20] Ligand_Design, User Guide, vers. 2.1. CA: Biosym Technologies, 1992.